Gaussian Processes for Survival Analysis

Imperial CSML Reading Group

Michael Komodromos

February 26, 2021

Table of Contents

Recap: Survival Analysis

Survival Analysis models time to failure events. Let T be a random variable modelling time to failure with pdf $f(t)$, then:

Survivor Function: $S(t) = 1 - F(t)$

Hazard Rate: $\lambda(t) = f(t)/S(t)$

Recap: Survival Analysis Identities

Some useful identities

$$\lambda(t) = \frac{f(t)}{S(t)} \tag{1}$$

$$= -\frac{S'(t)}{S(t)} \tag{2}$$

$$= -\frac{d}{dt}\log\left(S(t)\right) \tag{3}$$

Re-arranging Eq. (3) gives

$$S(t) = \exp\left(-\int_0^t \lambda(s)ds\right) \tag{4}$$

Recap: Gaussian Processes

Gaussian Processes: a collection of random variables, any finite number of which have a joint Gaussian dist.

A GP is specified by it's mean function $m(x) = \mathbb{E}[f(x)]$ and kernel function $k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]$.

We write

$$f(x) \sim GP(m(x), k(x, x')) \tag{5}$$

or just $f \sim GP$

Recap: Gaussian Processes

Suppose

$$y_i = f(X_i) + \epsilon_i, \tag{6}$$

where $\epsilon_i \sim N(0, \sigma^2)$.

We are interested in describing the uncertainty of f. Under a Bayesian framework we have

$$p(f|X, y) \propto p(f|X)p(y|f, X). \tag{7}$$

Recap: Gaussian Processes

$$p(f|X,y) \propto p(f|X)p(y|f,X). \tag{8}$$

If we place a GP prior over f, then we have

$$f|X \sim N(m(x), K(X,X)) \tag{9}$$

and

$$y|f,X \sim N(m(x), k(X,X) + \sigma^2 I_n) \tag{10}$$

Given our likelihood and prior are normally dist we know by conjugacy that $f|X,y \sim N$

Onto the paper!

A one slide overview

Semi-parametric method where the hazard rate

$$\lambda(t) \quad = \quad \underbrace{\lambda_0(t)}_{\text{parametric}} \quad \times \quad \underbrace{\sigma(l(t))}_{\text{non-parametric}} \tag{11}$$

where $\sigma(\cdot)$ is a link-function.

In other words, we specify the baseline hazard rate $\lambda_0(t)$ from a parametric distribution (Weibull, Exp etc...) and we model $l(\cdot)$ via a GP.

Model

We're interested in modelling time to failure events $T \in \mathbb{R}^+$ which has density $f(t)$, survivor function $S(t)$ and hazard rate $\lambda(t) = f/S$.

Model (without covariates)

$$l(\cdot) \sim GP(0, k), \quad \lambda(t)|l, \lambda_0 = \lambda_0(t)\sigma(l(t)), \quad T_i|\lambda \overset{iid}{\sim} f(t) \tag{12}$$

where $\sigma(x) = (1 + e^{-x})^{-1}$

Note

$$f(t) = \lambda(t) \exp\left(-\int_0^t \lambda(s) ds\right) \tag{13}$$

Interpretation

Let's look into $\lambda(t)$ a bit more. Under our model definition we have

$$\lambda(t) = \lambda_0(t)\sigma(l(t)) \tag{14}$$

So $\lambda_0(t)$ is derived from a parametric distribution and $\sigma(l(t))$ adjusts the baseline hazard rate by some multiplicative term. Noting that $0 \leq \sigma(x) \leq 1$.

Interpretation

Example

If we believe the baseline hazard rate is given by $1/\mu$ i.e. is the hazard rate of $\mathrm{Exp}\,(1/\mu)$ recalling

$$\frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} = \frac{1}{\mu} \frac{e^{-t/\mu}}{(1 - (1 - e^{-t/\mu}))} = \frac{1}{\mu}. \tag{15}$$

Then our model adjusts the baseline hazard by $\sigma(l(t))$ for some $t > 0$

Interpretation

Example

If we believe the baseline hazard rate is given by $\beta t^{\alpha-1}$ for $\alpha, \beta > 0$, the hazard rate of a Weibull dist, then we're adjusting the baseline hazard by $\sigma(l(t))$.

How much adjustment are we expecting?

Let's consider, $E[\sigma(X)]$ where $X \sim N(0,1)$.

Result: MacLaurin expansion of $\sigma(x)$

$$\sigma(x) = \frac{1}{2} + \frac{1}{4}x - \frac{1}{48}x^3 + \cdots \tag{16}$$

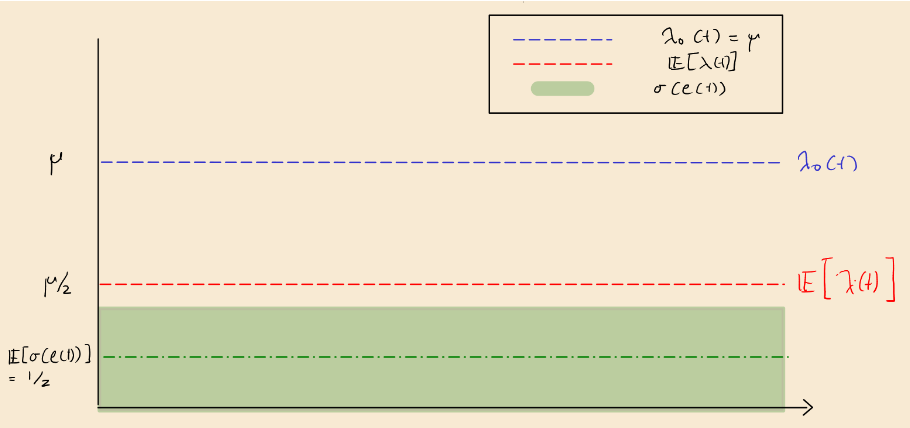
Now consider

$$\begin{aligned}
\mathbb{E}[\sigma(X)] &= \int_{\mathbb{R}} \sigma(x) f_X(x) dx && (17)\\
&= \int_{\mathbb{R}} \left(\frac{1}{2} + \frac{1}{4}x - \frac{1}{48}x^3 + \cdots \right) f_x(x) dx && (18)\\
&= \frac{1}{2} && (19)
\end{aligned}$$

Recall, $\mathbb{E}[(X-\mu)^p] = 0$ when p is odd.

Visualisation

Figure: Visualisation where we take the baseline hazard $\lambda_0(t) = \mu$



Making sure the model defines a well-behaved Survival function.

Proposition 1.

Let $(l(t))_{t\geq 0} \sim GP(0, k)$ be a stationary Gaussian Process. Suppose $k(s)$ is non-increasing and $\lim_{s\to\infty} = 0$. Further, assume there exists $K > 0$ and $\alpha > 0$ such that $\lambda_0(t) > Kt^{\alpha-1}$ for $t \geq 1$.

Then for random survival function $S(t)$ associated with $l(t)$ we have $\lim_{t\to\infty} S(t) \xrightarrow{p} 0$.

Where $\xrightarrow{p}$ denotes convergence in probability. A proof is given in the supplementary materials.

Any questions so far?

Adding Covariates

Covariates are introduced into the model through the kernel. Recall that we can construct kernels by performing basic operations (addition, multiplication) on multiple kernels.

Adding Covariates

Let $X \in \mathbb{R}^d$ be our covariates and t be the observed failure time, then for pairs (t, X) and (s, Y) we have

$$K((t, X), (s, Y)) = K_0(t, s) + \sum_{i=1}^{d} X_i Y_i K_i(t, s) \tag{20}$$

Model with Covariates

Assuming we have some covariates $X_i \in \mathbb{R}^d$, then the new model is given by

Model (with covariates)

$$l(\cdot) \sim GP(0, K), \quad \lambda_i(t)|l, \lambda_0(t), X_i = \lambda_0(t)\sigma(l(t, X_i))$$

$$T_i|\lambda \overset{\text{ind}}{\sim} \lambda(T_i)e^{-\int_0^{T_i} \lambda_i(s)ds}$$

We need to make sure of is that K is stationary

Inference

Recall, under our model we have

$$p(T_i|\lambda) = \lambda(T_i)e^{-\int_0^{T_i} \lambda_i(s)ds}. \tag{21}$$

In general λ_i is not analytically tractable as λ_i is defined by a Gaussian Process.

Numerical methods can be used to approximate $\int \lambda(t)dt$ but these are computationally expensive.

Instead a data-augmentation scheme based on **Poisson thinning** is used.

Experiments

Conclusions